数理解析研究所講究録
第 777 巻 1992 年 45-50

VALIDITY OF THE BOX-COX TRANSFORMATION OF THE EXPONENTIAL DISPERSION MODEL WITH POWER VARIANCE FUNCTION

広島大学 総合科学部　西井 龍映　(Ryuei NISHII)

1. Exponential Dispersion Model

指数型分布族に属する確率密度関数に集中度パラメーター $\lambda > 0$ を加えた族

$$f(x;\lambda,\theta) = a(x;\lambda)\exp[\lambda\{\theta x - \kappa(\theta)\}], \quad \theta \in \Theta, \quad \lambda \in \Lambda$$

を Exponential Dispersion (ED) Model という. $X \sim ED$ のとき、

$$E(X) = \kappa'(\theta),\ V(X) = \kappa''(\theta)/\lambda,\ \log E(e^{tX}) = \lambda\{\kappa(\theta + t/\lambda) - \kappa(\theta)\}$$

となる. このキュムラント母関数の形から、n 個の独立標本の平均 $\bar{X}$ は $f(x;n\lambda,\theta)$ に従うことがわかる。すなわち、λは n と同じ様な役割をする。そのため、中心極限定理により Xを平均、分散で基準化すると $\lambda \to \infty$ のとき N(0,1) に近づく.

さて、$\kappa(\theta)$ が

$$\kappa_\alpha(\theta) = (1 - 1/\alpha)\{-\theta/(1-\alpha)\}^\alpha\ (\alpha \neq 0), \quad \kappa_0(\theta) = -\log(-\theta)\ (\alpha = 0)$$

で与えられる時, 平均と分散はそれぞれ

$$E(X) = \{-\theta/(1-\alpha)\}^{\alpha-1}(=\mu), \qquad V(X) = \{-\theta/(1-\alpha)\}^{\alpha-2}/\lambda$$

となる. この分布族の分散関数 $\kappa_\alpha''(\theta)$ は平均の巾乗, すなわち $\kappa_\alpha''(\theta) = \mu^p$, $p = (\alpha-2)/(\alpha-1)$ という関係がある. この分布族を巾分散関数を持つ EDといい、$ED^{(\alpha)}$ で記す. ただし、$ED^{(\alpha)}$ は、$\alpha \leq 2, \alpha \neq 1$ のとき存在することが Jorgensen(1987) により示されている.

$ED^{(\alpha)}$ のサポート

$$\left(\begin{array}{ccc} \alpha = 2 & \text{正規分布} & (-\infty,\infty) \\ 1 < \alpha < 2 & & (-\infty,\infty) \\ 0 < \alpha < 1 & & (0,\infty) \\ \alpha = 0 & \text{ガンマ分布} & (0,\infty) \\ -\infty < \alpha < 0 & & [0,\infty) \\ \alpha = -\infty & \text{ポアソン分布} & \{0,1,\ldots,\} \end{array}\right)$$

$ED^{(\alpha)}$ の密度関数は $0 < \alpha < 1, 1 < \alpha < 2$ のときは安定分布の密度関数を利用して表現できる．$\alpha < 0$ の時はガンマ分布のポアソン分布による混合分布であるため，0 という値をとる確率は正で，正の範囲では連続型である．(Nishii, 1990, 参照)

以下 $0 < \alpha < 1$ の場合を考える．($\alpha = 1/2$ のときが逆ガウス分布)　このとき $ED^{(\alpha)}$ は正値分布族であり、密度 $f_\alpha(x;\lambda,\theta)$ は

$$(1) \quad f_\alpha(x;\lambda,\theta) = \gamma q(\gamma x;\alpha)\exp[\lambda\theta x - \lambda(1-1/\alpha)\{-\theta/(1-\alpha)\}^\alpha], \quad \gamma = \alpha^{1/\alpha}\{(1-\alpha)\lambda\}^{1-1/\alpha},$$

$$(2) \quad q(x;\alpha) = -\frac{1}{\pi x}\sum_{k=1}^{\infty}\frac{\Gamma(\alpha k+1)}{k!}(-x^{-\alpha})^k \sin(\pi\alpha k)\ (x > 0)\ ;\ = 0\ (x \le 0).$$

で与えられる．ここで $q(x;\alpha)$ は安定分布の密度関数である．

2．$ED^{(\alpha)}$の正規化変換

$X \sim ED^{(\alpha)}$ とすると，Xを平均と分散で基準化した統計量 $\sqrt{\lambda}(X-\mu)/\mu^{(\alpha-2)/(2\alpha-2)}$ は λ が大きくなると N(0,1)+$O_p(\lambda^{-1/2})$ となる．この収束の誤差のオーダーを $O_p(\lambda^{-1})$ へと加速する正規化変換は Box-Cox 変換：

$$X \longrightarrow X^{(1-2\alpha)/(3-3\alpha)}\ (\alpha \ne 1/2)\ ;\ X \longrightarrow \log X\ (\alpha = 1/2)$$

であることが知られている (Nishii, 1990)．たとえば

ガンマ分布	$\alpha = 0,$	$p = 2,$	の正規化変換 $X^{1/3}$,
逆ガウス分布	$\alpha = 1/2,$	$p = 3,$	の正規化変換 $\log X$,
正規分布	$\alpha = 2,$	$p = 0,$	の正規化変換 X,
ポアソン分布	$\alpha = -\infty,$	$p = 1,$	の正規化変換 $X^{2/3}$.

$ED^{(\alpha)}$は任意の次数のモーメントを持つため，正規化変換で得られる統計量の Edgeworth 展開は妥当 (Valid) である．ところがこの変換はオーダーの改良をするだけではなく，変換された確率変数は収束する Gram-Charier 展開（無限次の Edgeworth 展開）を持つようになる．すなわち、別の観点からも N(0,1) への収束を加速させていることを意味する。それを示すには次の Cramer による補助定理が重要な役割を果たす．

補助定理. (Cramer, 1925) Xは密度$f(x)$を持つ確率変数とし、密度の Gram-Charier 展開(G-C 展開)を：

$$\hat{f}(x)=\sum_{j=0}^{\infty}c_jH_j(x)\phi(x)/j!,\quad c_j=\int_{-\infty}^{\infty}H_j(x)f(x)dx$$

とおく. もし$f(x)$が任意の有限区間で有界変動で、$\lim_{x\to\pm\infty}f(x)=0$ かつ$\int_{-\infty}^{\infty}f(x)e^{x^2/4}dx$ が有界なら$\hat{f}(x)$ はいたるところ $\{f(x+0)+f(x-0)\}/2$ に収束し、有限区間では一様収束する. ただし、$\phi(x)$ は N(0,1) の密度関数、$H_j(x)$ はエルミート多項式である.

注意：G-C 展開は N(0,1) による展開であるから、密度関数$f(x)$は平均0, 分散1に近いようにしておいたほうが近似が良い。

3. $ED^{(\alpha)}$の巾変換と Gram-Charier 展開

Cramer の補助定理より 定理1，2を得る.

定理1. 確率変数Uはガンマ分布,$ED^{(0)}$, に従うとする. $0<\beta<1/2$ かつ $\lambda>\beta$ ならば, U^βをその漸近分散、漸近平均で基準化した

(3) $$\sqrt{\lambda}(U^\beta-\mu^\beta)/(\beta\mu^\beta)$$

の密度関数の G-C 展開は収束する.

(注意: $\beta=1/3$ のときが正規化変換, $\beta=0$ のときが分散安定化変換)

定理2. 確率変数Vは逆ガウス分布,$ED^{(1/2)}$, に従うとする。-1/2< $\beta<1/2$ を満たすβ に対し

(4) $$\sqrt{\lambda/\mu}(V^\beta-\mu^\beta)/(\beta\mu^\beta)\ (\beta\neq 0)\ ,\ \sqrt{\lambda/\mu}\log(V/\mu)\ (\beta=0)$$

の密度関数の G-C 展開は収束する.

(注意: $\beta=0$ すなわち log 変換が正規化変換, $\beta=$ -1/2 のときが分散安定化変換)

定理1、2の証明は確率変数(3),(4)の密度関数を求め、補助定理の十分条件を満たす事を示せばよい。

一般に $ED^{(\alpha)}$ $(0 < \alpha < 1)$ について同様の定理を示すためには、$f_\alpha(x;\lambda,\theta)$ のテイルにおける評価が必要となる。次の定理は原点に近いときの $f_\alpha(x;\lambda,\theta)$ の近似に有効である。

定理3. (2) で定義された安定分布の密度関数 $q(x;\alpha)$ のサドルポイント近似は

$$\hat{q}(x;\alpha) = \frac{x^{-(2-\alpha)/(2-2\alpha)}}{\sqrt{2\pi(1-\alpha)}} \exp\{-(1-\alpha)\alpha^{\alpha/(1-\alpha)}x^{-\alpha/(1-\alpha)}\}$$

であり, 近似誤差 $R(x;\alpha) = q(x;\alpha) - \hat{q}(x;\alpha)$ は不完全ガンマ関数を用いて評価できる.

定理3と(1),(2)により, 原点の近傍における $f_\alpha(x;\lambda,\theta)$ のサドルポイント近似とその近似誤差の評価が得られる. 他方、無限大の近傍における $f_\alpha(x;\lambda,\theta)$ は密度関数(1)に含まれる $e^{\lambda\theta x}$, $(\theta < 0)$ のオーダーで0に収束する。以上より

定理4. 確率変数 W は密度関数 $f_\alpha(x;\lambda,\theta)$ に従うとする. $-\alpha/(2-2\alpha) < \beta < 1/2$ ならば,

(5) $\sqrt{\lambda}(W^\beta - \mu^\beta)/(\beta\mu^{\beta-\alpha/(2\alpha-2)})$ $(\beta \neq 0)$, $\sqrt{\lambda}\log(W/\mu)/\mu^{-\alpha/(2\alpha-2)}$ $(\beta = 0)$

の G-C 展開は収束する.

注意1: $\beta = (1-2\alpha)/(3-3\alpha) \in (-\alpha/(2-2\alpha), 1/2)$ のときが正規化変換, $\beta = -\alpha/(2-2\alpha)$ のときが分散安定化変換)

注意2: 定理4で $\alpha = 1/2$ のおくと定理2が得られるが, $\alpha = 0$ とおいて定理1が得られるわけではない.

4. 数 値 例

$f_\alpha(x;\lambda,\theta)$ を98人の男子大学生の体重のデータに適応させる. 標本平均, 分散はそれぞれ 61.02Kg, 38.79Kg^{2}である.またパラメーターの MLE は $\hat{\alpha} = .878$, $\hat{\lambda} \doteq .679 * 10^{15}$, $\hat{\theta} = -.283 * 10^{-15}$ となる. Figure では,

$f_{\hat{\alpha}}(x;\hat{\lambda},\hat{\theta})$ (Exact.pdf),

そのサドルポイント法による近似密度関数 (Saddle.pdf),

定理 3 による近似密度の上限，下限 (Upper.bound, Lower.bound),

巾正規分布を当てはめたときの密度関数 (Box_Cox.transform)

をグラフで表した．サドルポイント近似や巾正規分布が $ED(\alpha)$ の密度関数によくあてはまっていることがわかる．またサドルポイント近似をしたときの上限、下限もほぼ満足すべき性能を示している。

謝 辞

Cramer(1925) の論文を探し出して下さった数理解析研究所及び京都大学理学部数学教室の図書室司書の方々に感謝の意を表します。

引 用 文 献

Cramer, H. (1925). On some classes of series used in mathematical statistics. Skandinaviske Matematikercongres, Copenghagen, 399-425.

Jorgensen, B. (1987). Exponential dispersion models (with discussion). J. R. Statist. Soc. B 49, 127-162.

Nishii, R. (1990). The relation between the normalizing Box-Cox transformations and the exponential dispersion models with power variance functions. Tech. Repo. #279, Statistical Research Group, Hiroshima University.

FIGURE : Saddlepoint Approximation of ED Model fitted to weight data